

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 02:51:12 ; Search time 1672.99 Seconds

(without alignments)  
13446.329 Million cell updates/sec

Title: US-09-768-781-2

Perfect score: 1389

Sequence: 1 atgaacaagaagaccacaaca.....caaggcaagtggtgtctga 1389

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: em\_estba.\*

2: em\_esthum.\*

3: em\_estin.\*

4: em\_estmu.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_estc.\*

9: gb\_esti.\*

10: gb\_est2.\*

11: gb\_hic.\*

12: gb\_est3.\*

13: gb\_est4.\*

14: gb\_est5.\*

15: em\_estfun.\*

16: em\_estom.\*

17: gb\_gss.\*

18: em\_gss\_hum.\*

19: em\_gss\_inv.\*

20: em\_gss\_pln.\*

21: em\_gss\_vrt.\*

22: em\_gss\_fun.\*

23: em\_gss\_mam.\*

24: em\_gss\_mus.\*

25: em\_gss\_other.\*

26: em\_gss\_pro.\*

27: em\_gss\_rod.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	303.4	21.8	1711	11 AK007734	AK007734 Mus muscu
2	294.6	21.2	1774	11 BC027097	BC027097 Mus muscu
3	259.4	18.7	624	10 BB200527	BB200527 BB200527
4	247	17.8	810	12 BG675952	BG675952 602622310
5	240.6	17.3	676	9 AL652777	AL652777 AL652777
6	238.4	17.2	392	10 BE031611	BE031611 130228 NA

7	206.2	14.8	881	13 BI827782	BI827782 603075458
8	198.4	14.4	201	9 AA256009	AA256009 2828606.r
9	193.4	13.9	678	9 AL638533	AL638533 AL638533
10	182.6	13.1	793	13 BI546216	BI546216 603188602
11	174.2	12.5	855	13 BI463670	BI463670 603207340
12	173.2	12.5	385	9 AA791818	AA791818 vs61c09.r
13	170.4	12.3	691	17 CNS02LFV	AL202756 Tetraodon
14	113.2	8.1	691	17 AG085155	AG085155 Pan trogl
15	109.2	7.9	666	13 BI461092	BI461092 603206970
16	106	7.6	1035	17 CNS01TL3	AL166656 Tetraodon
17	102	7.3	346	14 H87640	H87640 Yw17g11.r1
18	92.6	6.7	535	14 BM683941	BM683941 UI-E-EJ1-
19	92.6	6.7	537	14 BM931807	BM931807 UI-E-EJ1-
20	79.2	5.7	921	12 BF672674	BF672674 602152274
21	73.2	5.3	447	9 AL697050	AL697050 wc85c10.x
22	63.4	4.6	694	10 BB638337	BB638337 BB638337
23	61.2	4.4	869	17 CNS032YG	AL225457 Tetraodon
24	58.4	4.2	387	17 AQ087397	AQ087397 HS 2187.A
25	58.4	4.2	627	12 BF211581	BF211581 601812270
26	51	3.7	550	10 BB611251	BB611251 BB611251
27	50.2	3.6	857	17 CNS03K4E	AL247703 Tetraodon
28	49.8	3.6	574	17 AZ593746	AZ593746 IM0405N22
29	48.4	3.5	755	9 AL653785	AL653785 AL653785
30	47.8	3.4	930	14 BQ718536	BQ718536 AGENCOURT
31	45	3.2	431	10 AW374923	AW374923 MR0-CT006
32	45	3.2	464	9 AI923983	AI923983 wn49d11.x
33	45	3.2	588	9 AI831578	AI831578 wj39c11.x
34	45	3.2	588	12 BF195029	BF195029 7091h12.x
35	45	3.2	589	14 BM842656	BM842656 K-EST0120
36	45	3.2	609	10 AW966092	AW966092 EST378165
37	45	3.2	886	13 BI753850	BI753850 603027531
38	45	3.2	917	9 AL551219	AL551219 AL551219
39	45	3.2	922	12 BE735816	BE735816 601305125
40	45	3.2	926	12 BE735816	BE735816 601305125
41	45	3.2	993	13 BM472443	BM472443 AGENCOURT
42	45	3.2	1009	9 AL547473	AL547473 AL547473
43	45	3.2	1042	13 BM470447	BM470447 AGENCOURT
44	45	3.2	1149	13 BM466219	BM466219 AGENCOURT
45	44.8	3.2	877	13 BI463669	BI463669 603207339

## ALIGNMENTS

RESULT 1	AK007734	1711 bp	mRNA	linear	HTC 19-JAN-2002
LOCUS	AK007734	Mus musculus	10 day old male pancreas cDNA, RIKEN full-length		
DEFINITION	AK007734	enriched library, clone:1810038K19:McLeod syndrome gene homolog, full insert sequence.			
ACCESSION	AK007734	GI:12841469			
VERSION	AK007734.1	GI:12841469			
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (strain:C57BL/6J) 10 day old male pancreas cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library				
ORGANISM	Clone:1810038K19.				
	Mus musculus				
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
AUTHORS		High-efficiency full-length cDNA cloning			
TITLE		Meth. Enzymol. 303, 19-44 (1999)			
JOURNAL		99279253			
MEDLINE		10349636			
PUBMED					
REFERENCE	2	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
AUTHORS		Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes			
TITLE		Genome Res. 10 (10), 1617-1630 (2000)			
JOURNAL		20499374			
MEDLINE		11042159			
PUBMED					



```
QY 574 GGCTCAGTGGCCCCAGCTGACCTATCAGCTCTATGTGAGCCTGATCTCTGCAGAGGTTCCC 633
|||
Db 645 GGCTCTGCCCCACAACTGACCTCAACTGATATATAAATGCTTGGAGCAGAAATCACT 704
|||
QY 634 CTGGGTAGATGTGCTTAATGATATTTCCCTGGTATCTGTACCTATGGGGCCACCCCTT 693
|||
Db 705 ACTGGAAGATGCTTTCATCATGACCTTGTCCCTGTCTGCTATTTGTTATATGGCGCTTACGT 764
|||
QY 694 TGCAATATGTTGGGTATCCAGATCAAGTACATGACTACAGATTCGCTTGGGCCACTA 753
|||
Db 765 TGCAATATGTTGGGTATCCAGATCAAGTACATGACTACAGATTCGCTTGGGCCACTA 824
|||
QY 754 GAAGTCTCTGTCATCACCATCTGGCGGACATTTGGAGATCACTTCCCGCTCTCTGATCTG 813
|||
Db 825 GCTTATGCTGATCTCTCTCTGAGAGCTTTTGAGATTCACCCGAGTATCTGCTCTG 884
|||
QY 814 GTGCTCTCTCAGCCACTTTGAAATTTGAAGCTGTGCGCTTCTAGTGTCTCAACTTCTG 873
|||
Db 885 GTCTCTTTTACTCTGCTGAGATCTGGGTGGTGGCAGTCATCTGCTCAACTTCTTC 944
|||
QY 874 ATCATCTCTTTGAGCCTGCAATTAAGTTCTGGAGAGTGTGCGCAGATGCCCAATAAC 933
|||
Db 945 AGCTTCTTCTTATATCCCTGATCGTTTCTTGGTGCAGTGGCTCCCCATTCCTTGAGAAC 1004
|||
QY 934 ATTGAGAAATACTTTCAGCGGGTGGCACTCTGGTGGTCTGATTTTCACTCACCATCTC 993
|||
Db 1005 ATAGAGAGGCCCTTAAGTAGGGTGGTACCACTCTGCTCTGCTTCTCCTCACTTACTC 1064
|||
QY 994 TATGCTGGCATCAACTTCTTCTGCTGCTGAGCTTTGCACTTTGAGGTTGGCAGACAGAT 1053
|||
Db 1065 TATGCGGCATCAACATGTTCTGCTGGTCACTGTTTCACTGAAATCGACAATCCGGAG 1124
|||
QY 1054 CTGCTGACAAAGGGCAGAACTGGGACATATGGCCTGCACTATAGTGTGAGGTTGTA 1113
|||
Db 1125 CTATCAGCAAGTCCCAAGTGTGTACCGCTGCTCATTTACTACATGACGAGATTCATC 1184
|||
QY 1114 GAGATGTGATCATGCTGCTTGGTTTAAAGTTCTTGGAGTGAAGTCTTACTGAATTAAC 1173
|||
Db 1185 GAGAACTCGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1244
|||
QY 1174 TGTCATTCCTTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1233
|||
Db 1245 TGTCGCTCTGTTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1304
|||
QY 1234 CTCCTTTTCTTCCAGTACTTGCATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1286
|||
Db 1305 CTTGTGTTCTATCATGTTTTCACCCCTTGCACAAAGAGCTCTTCTCTCCTCAGTGT 1357
|||
```

```
RESULT 2
BC027097 1774 bp mRNA linear HTC 07-AUG-2002
LOCUS Mus musculus, McLeod syndrome gene homolog, clone IMAGE:5012421,
DEFINITION mRNA.
ACCESSION BC027097
VERSION BC027097.1 GI:20071516
KEYWORDS HTC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1774)
Strausberg, R.
Direct Submission
Submitted (04-APR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcaphs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
```

cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,

Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,

Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 45 Row: d Column: 3

This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 12963702  
This clone has the following problem: frame shifted.

#### FEATURES

##### Source

1. 1774

/organism="Mus musculus"

/db\_xref="taxon:10090"

/map="CZECH II"

/clone="IMAGE:5012421"

/tissue\_type="Mammary tumor metastasized to lung. Tumor

arose spontaneously from a senescent normal mammary

(clonal) outgrowth infected with the virus MMTV."

/clone\_lib="NCI CGAP\_Lu29"

/lab\_host="DH10B"

/note="Vector: pCMV-SPORT6"

BASE COUNT 386 a 513 c 441 g 434 t

##### ORIGIN

Query Match 21.2%; Score 294.6; DB 11; Length 1774;

Best Local Similarity 55.7%; Pred. No. 1.8e-63;

Matches 631; Conservative 0; Mismatches 489; Indels 13; Gaps 3;

QY 157 TCCACCTTTTGTACTGTGGGAGGCTGCATCTGCTTTGTACATCTGTAGATCTATCGA 216

Db 262 TCCGTGTTCTGTTGTCGGCGAGCGGGGGGCTCTACTTGAGCAGACCTACCGC 321

QY 217 AAGAAATAGTAACTTACTGGATGACATACACCTTTTCTTTTATGTTTTCATCAATT 276

Db 322 TCCGACGGGACCGCATGTGGCAGGTGTGACGCTGCTTCTCTCTGATGCGCTCGCC 381

QY 277 ATGTCAGTTGACCTTCATTTTGTCCACAGAGATCTAGCCAAAGATAAACCGCTATCA 336

Db 382 CTGTCAGTTTACGCTCTCTTCTGTCACCGCGACCTTCAGCCGCGATCGCCCACTGGCG 441

QY 337 TTATTTATGATCTAATCTCTTGGGACCTCTTATCAGATGTTTGGAGGCCATGATTAAAG 396

Db 442 CTGTCATGACCTTGCTCCAGCTCGGGCCCTGTACAGGTGTTGTAAGTCTTTTGTATC 501

QY 397 TACCTCAGCTGTGGAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 455

Db 502 TAC-----TGTCAGTCCGACCAAGATGAAGAACCTTATGTGAGCATCACTAAGAAA 552

QY 456 --GAAGATGCTAATAGATGGCGAGGAGGTGCTGATAGAAATGGAGGTGGGCGACCTCATC 513

Db 553 CGACAGATGCCAAAGATGCGCTTTTCAGAGGAGGTGGAGAAAGAGGTTGGCGAGCAGAA 612

QY 514 CGGACCTGGGTATGACCGCAATGCCCTACAAAGTATGTACAGATCCAGAGCTTCCCTG 573

Db 613 GGCAAGTGTATTACCCACCGTTCTGCATTCAGCCGGGATCAGTATACAGGCTTTCTCTA 672

QY 574 GGCTCAGTGGCCCCAGCTGACCTATCAGCTCTATGTAGCCCTGATCTCTGCAGAGGTTC 633

Db 673 GGCTCTGCCCCACAACTGACCTCAACTGATATATAAATGCTTGGAGCAGAAATCACT 732

QY 634 CTGGGTAGATGTGCTTAATGATTTTCCCTGGTATCTGTCACTATGGGGCCACCCCTT 693

Db 733 ACTGGAAGATGCTTTCATCATGACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 792

QY 694 TGCAATATGTTGGGTATCCAGATCAAGTACATGACTACAGATTCGCTTGGGCCACTA 753



```

QY 544 AACGTATGTCACAGATCCAGACCTTCTGGGCTCAGTGCCTCCAGTCCAGCTGACCTATCAGCTC 603
Db 181 AAAGTATGTCACAGATTCAGACCTTCTGGGCTCAGTGCCTCCAGTCCAGCTGACCTATCAGCTA 240

QY 604 TATGTGAGCTGTATCTGTGACAGAGTTCCCTGGGTAGAGTTGTG 648
Db 241 TATGTGAGCTGTATCTGTGACAGAGTCCCTGGGTAGAGGTGAG 285

RESULT 4
LOCUS BG675952 810 bp mRNA linear EST 01-MAY-2001
DEFINITION 602622310F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4747323 5',
mRNA sequence.
ACCESSION BG675952
VERSION BG675952.1, GI:13907348
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 810)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10597 row: c column: 04
High quality sequence start: 4
High quality sequence stop: 797.
FEATURES
Location/Qualifiers
source 1..810
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4747323"
/clone_lib="NCI_CGAP_Skn4"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (TI phage-resistant)"
/notes="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 214 a 178 c 169 g 249 t
ORIGIN
Query Match 17.8%; Score 247; DB 12; Length 810;
Best Local Similarity 99.6%; Pred. No. 1.5e-56;
Matches 258; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1131 CTTGGTTTTAAGTCTTTGGAGTGAAGTGTACTGAATTAAGTCTGATTCCTTTGATTGC 1190
Db 1 CTTGGTTTTAAGTCTTTGGAGTGAAGTGTACTGAATTAAGTCTGATTCCTTTGATTGC 60

QY 1191 CTTGCAGCTCATTTATCTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 1250
Db 61 CTTGCAGCTCATTTATCTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 120

QY 1251 CTTGCATCCATTGGCTCACTCTTCCACCAATAGTAGTACCTACCTCCATTGCTGCTG 1310
Db 121 CTTGCATCCATTGGCTCACTCTTCCACCAATAGTAGTAGTACCTCCATTGCTGCTG 180

QY 1311 CTGTCCAGCACCTTCGGACCGGTTGAGAACTCAGAGCCACCTTTGAGACTGGAAGC 1370
Db 181 CTGTCCAGCACCTTCGGACCA -GGTTGAGAACTCAGAGCCACCTTTGAGACTGGAAGC 239

```

```

QY 1371 AAGCAAGTGTGTGTCTGA 1389
Db 240 AAGCAAGTGTGTGTCTGA 258

RESULT 5
LOCUS AL652777 676 bp mRNA linear EST 13-DEC-2001
DEFINITION AL652777 XGC-gastrula Silurana tropicalis cDNA clone TGas029g15 5',
mRNA sequence.
ACCESSION AL652777
VERSION AL652777.1, GI:17663442
KEYWORDS EST.
SOURCE western clawed frog.
ORGANISM Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodidae; Silurana.
REFERENCE 1 (bases 1 to 676)
Huckle, E., Taylor, R., Ashurst, J.L., Zorn, A.M. and Rogers, J.
Sanger Xenopus tropicalis EST project 2001 (10_2001)
Unpublished (2001)
Contact: Huckle E
Sanger Centre
Hinxtion, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS SEQUENCE ID: TGas029g15.sp6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
FEATURES
Location/Qualifiers
source 1..676
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="TGas029g15"
/clone_lib="XGC-gastrula"
/dev_stage="gastrula (stages 10.5-13 mixed)"
/lab_host="Escherichia coli XL1-blue"
/notes="Vector: pCSI07; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from sug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCSI07 with EcoRI at the 5' end and NotI at the 3'
end."
BASE COUNT 138 a 159 c 155 g 224 t
ORIGIN
Query Match 17.3%; Score 240.6; DB 9; Length 676;
Best Local Similarity 59.9%; Pred. No. 8.1e-55;
Matches 402; Conservative 0; Mismatches 269; Indels 0; Gaps 0;

QY 643 GTTGTGCTAATGGTATTTTCCCTGGTATCTGTCACTATGGGGCCACCTTTGCAATATG 702
Db 6 GGTTGCTCTCATGGTATGTGTCTTGCCTCAACTACATATGTGCTTAATCCTAATGTC 65

QY 703 TTGGCTATCCAGATCAAGTACGATGACTACAAAGATTGCGCTTGGGCCACTAGAAATCCTC 762
Db 66 CTGGCCATACAAATTAAGTATGATGACTATAAAGTGTCTTAAGTGTGCCAGCGTCTCTG 125

QY 763 TGCATCACCATCTGGCGGACATTTGGAGATCACTTCCCGCTCCTGATTCCTGGTCTCTTC 822
Db 126 TGCATCTGATGTGGAGATGTCTGGAGATTGCTACCGGTGTACAGTTCTGGTCTCTTTT 185

QY 823 TCAGCCACTTTGAAATGAAGGCTGTGCCCTTCTAGTGCTCAACTTCTCTGATCATCTC 882
Db 186 TGTTCAGCTCTTAAGCGCTGGTGGCTGCAGTGGCTTAGCAAACTTACTGGCGCTCTTC 245

QY 883 TTGAGCCCTTGGATTAAGTTCTGGAGAAAGTGGTCCAGATGCCCAATPAACATTGAGAAA 942
Db 246 CTTTGTGCTTTGGGTTAAATCTTGAGCAGTGGGACAGCCTGCCCAAAACATAGAAAAG 305

QY 943 AACTTCAGCCGGGTGCGCACTCTGGTGTCTGTGATTTTCAGTCACCATCCTCTATGCTGCGC 1002

```

```

Db 306 AACTTTAGCTGGTTGGTACTGTGTGACAGCTCCCTGGGTGAGTAACCTTGTCTTTATTCAGCC 365
Qy 1003 ATCAACTTCTCTTGTGTGTGCTGAGCTTTGAGTGTGGCAGACAGAGATCTCGTCGAC 1062
Db 366 ATCAATATGTTTGTGTGTGCTGCTGTGCTGAGTAAACCTATCAGATCGGGACTTAATTCAC 425
Qy 1063 AAAGGGCAGAACTGGGGACATATGAGGCTGCACTATAGTGTGAGGTGGTAGAGAAATGTG 1122
Db 426 CAGTCACAAAACCTGGGGAGGTTAAGTCTTCAATTATACCAATTCGGTGTGCTCGAGAAATGCA 485
Qy 1123 ATCATGCTCTGGTTTAAAGTCTTTTGGAGTGAAGTGTACTGAATTAAGTCTCAATCC 1182
Db 486 GCCCTCATCTGCTCTGTGACTCTACAGGAGGAGTGTGTGAGTATTTCTGTTCCTCCCA 545
Qy 1183 TTGATTGCTTGGCAGCTCAATATTGCTTATCTGATTTCCATTGGCTTCATGCTCTCTTTTC 1242
Db 546 CTGCTGGTGGTTCAACTTTAGTAGATATGCCACCGCATCTCTTCATGCTGCTTTT 605
Qy 1243 TTCCAGTACTTGCATCCATTTGGCTCACTTTTACCCCAATATGATAGACTACCTCCAT 1302
Db 606 TACCAGTACCTTACCCCTTGGCGTCTCTTTTACGCCCAACAGTTGAAGACTGTTTGGCT 665
Qy 1303 TGTGCTGCTG 1313
Db 666 TGTGTTGCTG 676

RESULT 6
LOCUS BE031611 392 bp mRNA linear EST 09-JUL-2000
DEFINITION 130228 MRC IPiG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BE031611
VERSION BE031611.1 GI:8326620
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 392)
AUTHORS Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.,
and Keele, J.W.
TITLE Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCAGCAGC
Plate: 59 row: C column: 17
Seq primer: ATTAGGTGACACTATAG.
FEATURES
source
1..392
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC IPiG"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI; Library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos."
BASE COUNT 81 a 87 c 100 g 124 t
Query Match 17.2%; Score 238.4; DB 10; Length 392;

Db 306 AACTTTAGCTGGTTGGTACTGTGTGACAGCTCCCTGGGTGAGTAACCTTGTCTTTATTCAGCC 365
Qy 1003 ATCAACTTCTCTTGTGTGTGCTGAGCTTTGAGTGTGGCAGACAGAGATCTCGTCGAC 1062
Db 366 ATCAATATGTTTGTGTGTGCTGCTGTGCTGAGTAAACCTATCAGATCGGGACTTAATTCAC 425
Qy 1063 AAAGGGCAGAACTGGGGACATATGAGGCTGCACTATAGTGTGAGGTGGTAGAGAAATGTG 1122
Db 426 CAGTCACAAAACCTGGGGAGGTTAAGTCTTCAATTATACCAATTCGGTGTGCTCGAGAAATGCA 485
Qy 1123 ATCATGCTCTGGTTTAAAGTCTTTTGGAGTGAAGTGTACTGAATTAAGTCTCAATCC 1182
Db 486 GCCCTCATCTGCTCTGTGACTCTACAGGAGGAGTGTGTGAGTATTTCTGTTCCTCCCA 545
Qy 1183 TTGATTGCTTGGCAGCTCAATATTGCTTATCTGATTTCCATTGGCTTCATGCTCTCTTTTC 1242
Db 546 CTGCTGGTGGTTCAACTTTAGTAGATATGCCACCGCATCTCTTCATGCTGCTTTT 605
Qy 1243 TTCCAGTACTTGCATCCATTTGGCTCACTTTTACCCCAATATGATAGACTACCTCCAT 1302
Db 606 TACCAGTACCTTACCCCTTGGCGTCTCTTTTACGCCCAACAGTTGAAGACTGTTTGGCT 665
Qy 1303 TGTGCTGCTG 1313
Db 666 TGTGTTGCTG 676

RESULT 7
LOCUS BE031611 881 bp mRNA linear EST 04-OCT-2001
DEFINITION 603075458F1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5167070 5', mRNA sequence.
ACCESSION BE031611
VERSION BE031611.1 GI:15939332
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 881)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Prepared by: The I.M.A.G.E. Consortium (LLNL)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11415 row: d column: 15
High quality sequence stop: 782.
FEATURES
source
1..881
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_119"
/tissue_type="medulla"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH MGC Library."
BASE COUNT 257 a 163 c 194 g 266 t
Query Match 14.8%; Score 206.2; DB 13; Length 881;
Best Local Similarity 56.8%; Pred. No. 2.5e-45;

```





**AUTHORS** Huckle, E., Taylor, R., Ashurst, J. L., Zorn, A. M. and Rogers, J.  
**TITLE** Sanger Xenopus tropicalis EST project 2001 (10\_2001)  
**JOURNAL** Unpublished (2001)  
**COMMENT** Contact: Huckle E  
Sanger Centre  
Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: trop@sanger.ac.uk  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS SEQUENCE ID: TNeu018n12.sp6  
Sequencing primer: SP6  
This sequence is from a Xenopus Gene Collection (XGC) library  
constructed by Aaron M. Zorn.  
**FEATURES**  
    **source**  
        1..678  
            /organism="Silurana tropicalis"  
            /db\_xref="taxon:8364"  
            /clone\_lib="TNeu018n12"  
            /dev\_stage="neurula"  
            /lab\_host="Escherichia coli DH10B"  
            /note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA  
            was oligo dT primed from 5ug of poly A+ RNA from neurula.  
            EcoRI-NotI cut cDNA was then ligated into pCS107 with  
            EcoRI at the 5' end and NotI at the 3' end."  
**BASE COUNT** 225 a 156 c 153 g 144 t  
**ORIGIN**  
Query Match 13.9%; Score 193.4; DB 9; Length 678;  
Best Local Similarity 58.7%; Pred. No. 7.4e-42;  
Matches 335; Conservative 0; Mismatches 236; Indels 0; Gaps 0;  
Qy 642 AGTTGCTGTAATGATTTCCCTGGTATCTGTCACCTATGGGCGCCACCTTTGCAATAT 701  
Db 616 AGCTGTCTCATGGCTATGTGCTTCCTCAACTACATATGGTGCCTTAATCCTAATGT 557  
Qy 702 GTTGCTATCCAGATCAAGTACGATGACTACAAAGATTCGCCCTTGGGCCACTAGAAATCT 761  
Db 556 CTTGGCCATACAAATTAAGTATGATGACTATAAAGTTGCTAAGTGGCCAGGTTCT 497  
Qy 762 CTGATCACCATCTGGGACATTTGGAGTACATCTCCGCCCTCTGATTTCTGGTCTT 821  
Db 496 GTGCATCGTGTGGAGATGCTGGAGATGCTACCCGTGTACAGATTTCTGGTTCTTT 437  
Qy 822 CTCAGCCACTTTGAATTAAGGCTGTGCCCTTCTAGTGTCAACTTCTGATCATCTCT 881  
Db 436 TTGTTACGCTCTTAAGCCCTGGGTGCTGCTGAGTGGCTTAGCAAACTTACTGGCGCTTT 377  
Qy 882 CTTTGAGCCCTGGATTAAGTCTGGAGAAAGTGTGCCAGATGCCCAATAACATTGAGAA 941  
Db 376 CTTTGTGCTTGGTTAAATTTCTGGAGCAGTGGGACAGCCTGGCCCAAAACATAGAAA 317  
Qy 942 AAATTTAGCCGGGTCGGCACTCTGTGTGCTCTGATTTTCACTCACCATCTCTATGCTGG 1001  
Db 316 GAACTTTAGCTGTTTGGTACTGTGACAGTCTTGGGTGAGTAACTTCTTTTATTCAGC 257  
Qy 1002 CATCAACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1061  
Db 256 CATCAATATGTTTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 197  
Qy 1062 CAAGGGCAGAACTGGGACATATGGCCCTGCACATATAGTGTGAGGTGGTATAGAAATGT 1121  
Db 196 CCAGTCACAAAACCTGGGAGGATTAAGTCTTATATACCATTCGTTGCTCGAGATGC 137  
Qy 1122 GATCATGCTTTGGTTTAAAGTCTTTTGGAGTGAAGTGTATCTGAATTAATCTGCTATTC 1181  
Db 136 AGCCCTCATCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 77  
Qy 1182 CTTGATGCTTGGCAGCTCATTAATGCTTAT 1212  
Db 76 ACTGCTGGTGGTTCAACTCTTAGTAGGATAT 46  
**RESULT** 10

BI546216  
**LOCUS** 603188602P1 NIH\_MGC\_95 Homo sapiens cDNA clone IMAGE:5259961 5',  
**DEFINITION** mRNA sequence.  
**ACCESSION** BI546216  
**VERSION** BI546216.1 GI:15433528  
**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 793)  
NIH-MGC <http://mgc.ncl.nih.gov/>.  
**AUTHORS** National Institutes of Health, Mammalian Gene Collection (MGC)  
**JOURNAL** Unpublished (1999)  
**COMMENT** Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LHAM11655 row: c column: 02  
High quality sequence stop: 707.  
**FEATURES**  
    **source**  
        1..793  
            /organism="Homo sapiens"  
            /db\_xref="taxon:9606"  
            /clone="IMAGE:5259961"  
            /clone\_lib="NIH\_MGC\_95"  
            /tissue\_type="hippocampus"  
            /lab\_host="DH10B"  
            /note="Organ: brain; Vector: pBluescriptR (modified  
            pBluescript KG+); Site 1: BamHI; Site 2: SalI-XhoI (gcgag  
            ); Oligo-dT primed using primer 5'-TTTCTTTTCTTTTCTTN-3',  
            size-selected for average insert size 2.5 kb and  
            normalized to ROT 5. This is a primary library enriched  
            for full-length clones and constructed using the  
            Cap-trapper method (Carninci, in preparation). Library  
            constructed by M. Brownstein (NIH/NHGRI, National  
            Institutes of Health). Note: this is a NIH\_MGC Library."  
**BASE COUNT** 180 a 215 c 188 g 210 t  
**ORIGIN**  
Query Match 13.1%; Score 182.6; DB 13; Length 793;  
Best Local Similarity 56.9%; Pred. No. 7.2e-39;  
Matches 419; Conservative 0; Mismatches 304; Indels 14; Gaps 4;  
Qy 289 ACCCTCATTTTGTCCACAGATCTAGCCAAAGATAAACCGCTATCATTTATGTCAT 348  
Db 1 AGCGGCTCTTCTGTCACCGGACCTCAGCGCGACCGCCGCTCTGCTGCTGTCAC 60  
Qy 349 CTAATCCTCTTGGGACCTGTTATCAGATGTTTGGAGGCCATGATTAAGTACCTCACACTG 408  
Db 61 CTGCTGCAACTTGGGCGCGTTTTCAGGTGTTTGAAGTCTTCTGTCATCTAC----- 111  
Qy 409 TGGAAAGAGAGGAGGAGGAGGAGGCGCTATGTGAGCTCACCCGAAAGAG---ATGCTA 465  
Db 112 TTTTCAGTCAGGCAACATGAGAGGCTTATGTGTCAGTATCACCAGAGAGGCAATGCCA 171  
Qy 466 ATAGATGGCAGGAGGTGCTGATAGAAATGGGAGGTGGGCACTCCATCCGAGACCTGGCT 525  
Db 172 AAAAATGGCTCTCAGAGGAGATTGAGAAGGAGGTGGGCGCAGCAGAGCAAACTAATC 231  
Qy 526 ATGACCCCAATGCCCTACAAAGTATGTACAGATCCAAAGCCCTTCTCTGGCTCAGTGCCC 585  
Db 232 ACCCAACCATCATCGCTTCAGCCGGCGTCCGTTGATCCAGGCTTCTTGGGCTCAGCCCC 291  
Qy 586 CAGCTGACCTATCAGCTCTATGTGAGCCTGATCTCTGTCAGAGGTTCCCTCTGGGTAGATT 645



Db 292 CAGCTGACCCCTACAGCTGTACATATAGTGTATGACGACGAGAGCTCACTGTGTGAAGAAGT 351  
 QY 646 GTGCTAAATGGTATTTTCCCTGTATCTGTACCTATATGGGGCCACCTTTGCAATATGTTG 705  
 Db 352 CTCCTCATGACCATATCCCTGTGTCCATTTGTATGAGAGCCCTTGGCGTGCACATCTTA 411  
 QY 706 GCTATCCAGATCAAGTACGATGACTAC-AAGATTGCGCTTGGCCACCTAGAAAGTCTCTG 764  
 Db 412 CCGATCAAAATCAAGTACGATGAGTATGAAGTCAAGAGTGAAGGCTCTGGCCATGTCTG 471  
 QY 765 CATCACCATCTGGGGGACATTCGAGATCACTCCCGCCCTCTGATCTGGTCTCTCTTC 824  
 Db 472 TATCTCTCTGGAGGAGCTTTGAGATTGCCACTCGAGTTGTAGTCTGGTCTCTCTTTAC 531  
 QY 825 AGCCACTTTGAAATGAAGGCTGTGCCCTTCTAGTGTCTCACTTCCCTGATCATCTCTT 884  
 Db 532 CTCGCTCTGAAGACCTGGCGGTGGTTATATATCATCACTTCTTCAGTTCTTTCTTA 591  
 QY 885 TGAGCCCTGGATTAAATGTTCTGGAGAGTGGTCCGAGATGCCCAATAACATTCAGAAAAA 944  
 Db 592 GTACCCCTGGATCCTCTTCTGGTGGTGGTGGTCCCTCCCTCGAGACATAGAG-AAGG 650  
 QY 945 CTTGAGCCGGTGGGACATCTGGTGGTCTGATTTTCAGTCACTCATCTCTATGCTGGCAT 1004  
 Db 651 CCTCAGTAGAGTGGGACCACTGATGATGATGATGATGATGATGATGATGATGATGATG 710  
 QY 1005 CAACCTCTCTGCTGGT 1021  
 Db 711 TTAACCATGTTCTCTGCT 727  
  
 RESULT 11  
 BI463670  
 LOCUS 603207340F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:5273312 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BI463670  
 VERSION BI463670.1 GI:15254326  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 855)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L14M11689 row: o column: 09  
 High quality sequence stop: 826.  
 FEATURES  
 source  
 1..855  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5273312"  
 /clone\_lib="NIH\_MGC\_97"  
 /lab\_hosts="DH10B"  
 /note="Organ: testis; Vector: pBluescriptR (modified  
 pBluescript KS+); Site: BamHI; Site 2: SalI-XhoI (gtcggag  
 ); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
 size-selected for average insert size 2.2 kb and  
 normalized to ROT 5. This is a primary library enriched  
 for full-length clones and constructed using the  
 Cap-trapper method (Carninci, in preparation). Library

constructed by M. Brownstein (NIMH/NHGRI, National  
 Institutes of Health). Note: this is a NIH\_MGC Library."  
 BASE COUNT 257 a 156 c 186 g 256 t  
 ORIGIN

Query Match 12.5%; Score 174.2; DB 13; Length 855;  
 Best Local Similarity 57.0%; Pred. No. 1.5e-36;  
 Matches 347; Conservative 0; Mismatches 248; Indels 14; Gaps 1;  
  
 QY 8 CAGACCCACCAATTCAGAAAGAACTCGACATATGAGAGAGTTTATGAAATTCCTTGAGG 67  
 Db 72 CAAGGAAGCGAGAAAAATAAGCACCCCTGAGAAATGAGACAGTGTGTTGAAGAGATGGATG 131  
 QY 68 AGCCAAATGTGGATCGGGTTTCATCTCTGAGAGAGATGTCATCCGTGGAGCCCAACCCCC 127  
 Db 132 AGAAGACACAGGAGAGTTTCATCTTCGAAGAGAGAAATAGTCTTGGCCAGAGACTCC 191  
 QY 128 GATTATCTTTCCATTAGCATCTCTTTTCTCCACCTTTTGTACTGTGGGAGGCTGCAT 187  
 Db 192 ATCTAAGCTTTCTCTTTAGCATTTCTCTCAACTGTTCTTACTGTGGTGAAGTTGCCT 251  
 QY 198 CTGCTTTGTACATGTTAGAAATCTATCGAAGAAATAGTGAACACTTACTGGATGACATACA 247  
 Db 252 TTGGTTTATACATGTTTGAATTTTATCGAAAGCTAATGACACATTTCTGGATGTCATTTA 311  
 QY 248 CTTTCTTTCTTTATGTTTTCATCCATTTATGTCAGTTGACCCCTCATTTTGTCCACA 307  
 Db 312 CCATCAGCTTTTATTTGTTGGGGCAATTTTGGATCAAAATATCTCTGATGTTTCAACA 371  
 QY 308 GAGACTTAGCCAAAGATAAACCGCTTATCATTTATTTATGATCTTAATCTCTTGGGACCTG 367  
 Db 372 AAGACTTGAGGAGAAATAAGGCTGCATTTCTTTTGGCACATTTCTTTTAGGACCTA 431  
 QY 368 TTATCAGATGTTTGAGGCCATGATTAAGTACTCTACACTGTGGAAGAAAGAGGAGGAGG 427  
 Db 432 TTGTGAGGTGTTTGACACACCATTAGAAATTTACCAAAATGGTTGAAAAATCTTTAAACAGG 491  
 QY 428 AGGAGCCCTATGTCAGCTCCACCCGAAAGAGATGCTATAGATGCGGAGGAGGTGCTGA 487  
 Db 492 AGAAGGAAGA-----GACTCAAGTTAGCATCAAAAGAGAAACACGATGCT 537  
 QY 488 TAGAATGGAGGTGGGCCACATCCATCCGGACCCCTGGCTATGACCCCAATGCTCAAAAC 547  
 Db 538 GGCAAGGGAGATTGCATTCATCCGGGATAATTTCATGACAGAGAGGCTTTCAAGT 597  
 QY 548 GTATCTCAGATCCAAAGCCTTCTTGGGCTCAGTCCCGAGCTGACCTATCAGCTCTATG 607  
 Db 598 ACATGTCAGTGATTCAGGCTTTCTCGGTCTCTGTTCCCAATTAATTTTGCAGATGTATA 657  
 QY 608 TGAGCCTGA 616  
 Db 658 TCAGTCTCA 666  
  
 RESULT 12  
 AA791818  
 LOCUS  
 DEFINITION v61c09.r1 Stratiagene mouse skin (#937313) Mus musculus cDNA clone  
 IMAGE:1150768 5', similar to SW:HK\_HUMAN P51811 MEMBRANE TRANSPORT  
 PROTEIN XK ;, mRNA sequence.  
 ACCESSION AA791818  
 VERSION AA791818.1 GI:2854773  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 385)  
 AUTHORS Matra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.



```

DEFINITION Pan troglodytes DNA, clone: PTB-083B22.F, genomic survey sequence.
ACCESSION AG085155
VERSION AG085155.1 GI:16636957
KEYWORDS GSS.
SOURCE BAC Library clone: PTB-083B22.F.
ORGANISM Pan troglodytes
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominidae; Pan.
TITLE Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
JOURNAL Totoki, Y., Watanabe, H. and Sakaki, Y.
REFERENCE BAC end sequences of Library PTB
AUTHORS 2 (bases 1 to 691)
TITLE Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
JOURNAL Direct Submission
AUTHORS Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
JOURNAL and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimpbes@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI
FEATURES
Location/Qualifiers
source 1..691
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-083B22.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 230 a 151 c 127 g 181 t 2 others
ORIGIN
Query Match 8.1%; Score 113.2; DB 17; Length 691;
Best Local Similarity 60.9%; Pred. No. 7.7e-20;
Matches 184; Conservative 0; Mismatches 118; Indels 0; Gaps 0;
QY 640 AGAGTTGTGCTAATGTTATTTTCCTGTGTCTGTACCTATGGGGCCACCTTTGCAAT 699
DB 354 ATAGCATGTGTGATGACATGTTCCCTGTTATCAGTTACTTATGGGGCCATTCGCTGCAAT 295
QY 700 ATGTTGGCTATCCAGATCAAGTACAGTACGATCAAGATTCGCTTTGGGCCACTAGAAGTC 759
DB 294 ATACTGGCCATCCAGATCAGCAATGATGATGATACCTACCCTAAGCTACCGCGATGAAATTC 235
QY 760 CTCTGATACACCATCTCGGACATTTGAGATGATCACTCCCGCCCTCTGATTTCTGGTGCTC 819
DB 234 TTCTGTGTGATGTTGGCGTTTGTGGAGGTTATCTCACGTGTAGTACCTCTGGCAATTT 175
QY 820 TTCTCAGCCACTTTGAATTTGAAGGCTGTGCGCTTCTAGTGTCTCAACTTCTCTGATCATC 879
DB 174 TTCAATGCATCTCTGAAACTGAAAGGCTACCCCTTTTGTGTTAATCATATATTTTGTGTCA 115
QY 880 CTCTTTGAGCCCTGGATTAAGTTCTCGGAGAGTGTGTCGCCAGATGCCCAATAACATTTGAG 939
DB 114 TTGTTGCAACCGTGGCTGGAGTTTGGAAAGTGGAGCTCCGCGACGTCGTCTGTCGCG 55
QY 940 AA 941
DB 54 AA 53
RESULT 15

```

```

BI461092
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

BI461092 666 bp mRNA linear EST 21-AUG-2001
603206970F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5272840 5',
mRNA sequence.
BI461092
BI461092.1 GI:15251748
human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-x@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1688 row: k column: 17
High quality sequence stop: 663.
Location/Qualifiers
source 1..666
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5272840"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 125 a 212 c 197 g 132 t
ORIGIN

```

```

FEATURES
source 1..666
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5272840"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 125 a 212 c 197 g 132 t
ORIGIN

```

```

Query Match 7.9%; Score 109.2; DB 13; Length 666;
Best Local Similarity 53.7%; Pred. No. 9.5e-19;
Matches 278; Conservative 0; Mismatches 228; Indels 12; Gaps 2;
QY 157 TCCACCTTTTGTACTGTGGGAGGCTGCATCTGCTTTGTATCATGTTAGATCTATCGA 216
DB 156 TCCGTGTTCTCTGTTGTCGGCGAGACACGCGGCTCAGCTGAGCAGCACTACCGC 215
QY 217 AAGATAGTGAACCTTACTGGATGACATACACCTTTCTCTTTATGTTTTCATCCATT 276
DB 216 TCGGGCGGGACCGCATGTGGCAGCGCTGACGTGCTTTCTCGCTACTGCTTGGCGG 275
QY 277 ATGTCTCAGTTGACCCCTCATTTTGTCCACAGAGATCTAGCCAAAGATAAACCGCTATCA 336
DB 276 CTGTGTCAGCTCAGCTTCTCTTGTACCCGACCTCAGCCGCGACCGCCCGCTCGTA 335
QY 337 TTATTTATGTCATCTAATCTCTCTTGGGACCTGTTATCAGATGTTTGGAGCCATGATTAG 396
DB 336 CTGCTGTGTCACCTGCTGCAACTTGGGCCCTTTTCAGTGTGTTTGAAGTCTTCTGCATC 395
QY 397 TACCTCAGCTGTGGAAAGAGAGAGCAGGAGGAGCCCTATGTGAGCCTCAACCCGAAG 456
DB 396 TACTTTC-----AGTCAGGCAACAATGAAGAGCCTTATGTGATATCACCAGAG 446
QY 457 A---AGATGCTAATAGATGGCAGGAGGTGCTGATAGATGGAGGTGGCCACTCCATC 513
DB 447 AGGCAAAATGCCAAAAAATGGCCTCTCAGAGAGAGATTGAGAGAGGTGGGCCAGCAGAA 506

```

Qy	514	CGGACCTGGCTATGCACCGCAATGCCTACAAACGTATGTCACAGATCCAAAGCTTCCTG	573
Db	507	GGCAAACTAATCACCACCGATCAGCGTTACGCCGGCGTCGGTGATCCAGGCTTTCCTG	566
Qy	574	GGCTCAGTGCCCCCAGCTGACCTATCAGCTCTATGTGAGCCTGATCTCTGCAGAGGTTCCC	633
Db	567	GGCTCAGCCCCCAGCTGACCTACAGCTGTACATAAGTGTATGCAGCAGGACGTCACT	626
Qy	634	CTGGGTAGAGTTGTGCTAATGGTATTTTCCCTGGTATC	671
Db	627	GTGGAAGAAGTCTCCTCATGACCATATCCCTGTTGTC	664

Search completed: March 10, 2003, 05:49:12  
Job time : 1679.99 secs